IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

# AI IMAGE GENERATION WITH DALL-E AND STABLE DIFFUSION: A SURVEY

**[1]Rahul Sharma, [2]Harsh Jaiswal, [3]Nakul Singh Jadon, [4]Charul Bapna**

[1,2,3] Final Year B.Tech, Poornima Group of Institutions, Jaipur, Rajasthan, India
[4] Assistant Professor (AI & DS), Poornima Institute of Engineering and Technology, Jaipur, Rajasthan, India

***Abstract:*** Image generation has advanced significantly because of recent advances in artificial intelligence (AI), allowing machines to produce realistic photographs that nearly match those taken by humans. A thorough overview of AI image creation models is given in this survey, with an emphasis on the models' architectures, training methods, and applications. We categorize these models, along with their extensions and modifications, we categorize these models as the Stable Diffusion model and the DALL-E model. We go over the fundamental ideas of each model class and highlight their salient features and competencies. In addition, we offer a comparison of these models according to their scalability and computational complexity. Furthermore, we investigate the uses of image generation in a variety of industries, such as entertainment, fashion, content creation and design for demonstrating the usefulness and promise of these models in real-world contexts. Lastly, we analyze the drawbacks and limitations of the models currently used by AI to generate images and suggest future lines of inquiry to resolve these problems and improve AI's ability to produce realistic images.

***Keywords*** **- AI, Image generation, Generative, DALL-E, Diffusion, Deep learning, Optimization, Virtual, Unsupervised learning, text-to-image generation**

## INTRODUCTION

Image generation is the technique of utilizing artificial intelligence algorithms to make images that are indistinguishable from those created by people. This technique has found uses in a variety of fields, including art, design, entertainment, and practical problem resolution. AI models may create graphics ranging from photorealistic renderings to completely unique and bizarre views by recognizing the underlying patterns and structures inside datasets.

Image generation has emerged as a transformative field within artificial intelligence, revolutionizing our ability to generate visual information autonomously. AI systems can now generate amazingly realistic and imaginative visuals by combining advanced machine learning algorithms, neural networks, and significant computational power. In this survey, we examine the significance of two pioneering technologies in this field, DALL-E and Stable Diffusion.

## WORKING-OF-IMAGE-GENERATION-MODELS

### a) How DALL-E works?

DALL-E by OpenAI works on unsupervised learning. The model is trained on a large amount of text-image pair data, such as image caption data, and its parameters are improved through an optimization process. This optimization process is essentially a feedback loop in which the model predicts an output, compares it to the actual result, calculates the error, and then modifies the model parameters to reduce the error. This is accomplished using backpropagation and an optimization approach such as stochastic gradient descent. With appropriate data and training examples, DALL-E has established a remarkable ability to generate wholly new visuals that fit given verbal descriptions, including those that depict strange or previously encountered notions.

The combination of text and image data allows DALL-E to 'imagine' and generate visuals that are both contextually relevant to the input text and creatively original, much like a human artist would interpret a literary description. DALL-E is currently being used for a variety of purposes, including creating distinctive artworks and improving visual communication. For example, DALL-E can design a distinctive logo based on a given description or assist instructors by offering visual aids for abstract subjects.

Figure 1: Prompt - Cat with panda.
(via DALL-E AI)

Figure 2: Prompt - Super car on racetrack.
(via DALL-E AI)

**b) How Stable Diffusion works?**

StabilityAI's Stable Diffusion is an open-source generative AI model based on the diffusion deep learning paradigm. More specifically, they are generative models, which means they are trained to generate new data based on previously learnt information. As a side note, generative modeling is a sort of unsupervised learning that focuses on automatically detecting and learning patterns in input data such that the model can be used to generate new examples based on the original data. Stable diffusion received the moniker "diffusion" because its mathematical formulation is very similar to that of diffusion in physics. Stable Diffusion can be applied to text-to-image, image-to-image, image editing, and text-to-video generation. It can easily be tuned through local environment with a high performing GPU. It has many models that are trained on various datasets that provide variety in different image parameter in image generation.

Figure 3: Prompt - Castle in night surrounded by mountains. (via Stable Diffusion v5)

Figure 4: Prompt - Scuba diver underwater exploring coral reef. (via Stable Diffusion v5)

**APPLICATIONS OF IMAGE GENERATION**

- **Art and Designing:** As per user's requirements AI can generate unique artworks, generally Artists use AI to explore more creative alternatives and concept arts.
- **Education:** Textbook content can be turned into interactive and simulative stories, improving the teaching, and learning experience.
- **Content Creation:** Create visually appealing content such as website backgrounds, thumbnails, and photos depending on certain scenarios in a movie.
- **Architecture:** AI can be used to develop interior designs, project ideas, and virtual tours for prospective projects.

- **Film and Game Production:** Virtual landscapes for movies and games, realistic special effects (VFX), and character concept art for costume design.

- **Fashion:** Create photos of new apparel and accessories for fashion designers to use in their next creations, as well as an AI-enabled virtual presentation of products via AR/VR.

## LIMITAIONS OF IMAGE-GENERATION

While remarkable, image generation has limitations. One important difficulty is the model's tendency to produce images that lack actual creativity, frequently reproducing patterns detected in training data rather than providing truly novel content. Furthermore, the quality and realism of generated photos might vary, with AI struggling to capture the finer details and nuances present in human-created photographs. Furthermore, models such as DALL-E and Stable Diffusion require vast amounts of diverse and high-quality data to learn successfully, and their outputs may be constrained by biases in the training data. Controlling and comprehending the output of AI picture generating models can be difficult, and the computational resources needed for training and maintenance are significant. On the other side, the lack of AI regulations and laws in many countries is a major concern for controlling the spread of misinformation via image generation. Despite these limits, continued improvements, and research in this field show promise for resolving these issues, along with improved regulations by government authorities to avoid misuse of this technology.

## FUTURE SCOPE

The future of image generation promises a fascinating blend of creative exploration and transformative practical applications. As AI technologies continue to advance, we can anticipate an era where AI-generated images become more refined, versatile, and integrated into various aspects of our lives. In this discussion, we will explore the exciting possibilities and evolving landscape of AI-powered image generation.

Future development in fields of image generation:

**5.1 Virtual Environments:** Generative models could be incorporated into virtual reality (VR) and augmented reality (AR) applications, enabling the dynamic production of realistic environments and objects in real time.

**5.2 Conversational AI:** Image generating models could be included into conversational AI systems, allowing users to create images using natural language descriptions or dialogues.

**5.3 Metaverse and Virtual Worlds:** As virtual and augmented reality settings gain popularity, image generating models may play an important role in developing realistic, dynamic, and immersive virtual worlds and experiences.

**5.4 Cultural Heritage Preservation:** These models could be used to recreate or restore damaged or lost cultural items, historical places, or pieces of art, preserving them digitally for future generations.

**5.5 Multimodal Image Generation:** Future models may be able to generate images that include many modalities, such as visual and textual features, or images that include audio or other sensory data.

## CONCLUSION

Image generation has emerged as a transformational field in artificial intelligence, allowing machines to produce astonishingly accurate and inventive images. Models like DALL-E and Stable Diffusion have proven outstanding ability to generate graphics from written descriptions or other input data. However, technology has limits, including difficulties in creating original and creative content, potential biases from training data, and the significant computational resources necessary.

Despite these obstacles, the future of AI image production promises great opportunities. Image generation models may play an important part in producing immersive and dynamic digital experiences as technologies such as virtual and augmented reality, conversational AI, and the metaverse advance. Additionally, the potential applications in fields like cultural heritage preservation, multimodal content generation, and scientific visualization are promising.

As research in this field progresses, it is critical to address ethical concerns and potential abuses of this technology, such as the creation of deepfakes or false information. Responsible development, suitable laws, and interdisciplinary collaborations will be critical in maximizing the benefits of AI picture production while reducing its perils.

**REFERENCES**

[1] C. Ambrosio, "Unsettling robots and the future of art," Science (1979), vol. 365, no. 6448, pp. 38–39, Jul. 2019, https://www.science.org/doi/10.1126/science.aay1956/ (accessed Mar. 19, 2024).

[2] D. Carter, "Ai-Da creates the first portrait of the Queen painted by a humanoid robot | Creative Boom," Creative Boom, Jun. 01, 2022. https://www.creativeboom.com/news/ai-da-queen/ (accessed Mar. 19, 2024).

[3] J. Lewis, "Create art the AI way," Mid. Day, Sep. 25, 2022. https://www.mid-day.com/sunday-mid-day/article/create-art-the-ai-way-23247382 (accessed Mar. 19, 2024).

[4] A. Borji, "Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and DALL-E 2," ArXiv, Oct. 2022, Accessed: Dec. 10, 2022. [Online]. Available: https://www.researchgate.net/publication/364126542_Generated_Faces_in_the_Wild_Quantitative_Comparison_of_Stable_Diffusion_Midjourney_and_DALL-E_2 (accessed Mar. 19, 2024).

[5] H. Hassanzadeh, "How MidJourney And DALL·E 2 Help Designers to Create Unique Concepts?" Parametric Architecture, Aug. 15, 2022. https://parametric-architecture.com/how-midjourney-and-dalle-2-help-designers-to-create-unique-concepts/ (accessed Mar. 19, 2024).

[6] Z. Huang, "Analysis of Text-to-Image AI Generators," United States, 2022. [Online]. Available: https://digital.kenyon.edu/dh_iphs_ai/33/ (accessed Mar. 19, 2024).